Nondestructive Reading and Refreshment Circuit for Memristor-based Neuromorphic Synapse

A. H. Hassan
ONE Lab
Electronics and Communications
Engineering Department, Cairo
University, Giza 12613, Egypt.
ahhamed93@hotmail.com

Hassan Mostafa
Electronics and Communications
Engineering Department, Cairo
University, Giza 12613, Egypt.
Nanotechnology Department at Zewail
City for Science and Technology,
Cairo, Egypt.
hmostafa@uwaterloo.ca

Abstract*—Memristor, the fourth circuit element found ten years ago, has attracted most scientific and engineering community pioneers. Its proficiency in nanoscale devices makes it very useful in todays new fast circuit applications. Neuromorphic computing is the new processing technique which is brain-inspired rather than old von-Neumann processors. The main advantage of using memristor in neuromorphic is that it contains both the processing and storing blocks in one place. The main components of any neural networks are the neuron and synapse. In this paper, a recent memristor-based synapse used in literature is introduced with rather a nondestructive reading mechanism for the stored weight in the memristor.

Keywords—Memristive Systems, Neuromorphic Synapse, Nondestructive Reading, Refreshment control circuit.

I. INTRODUCTION

Neural Networks, NNs, is the concern of engineers nowadays because of their fast computation capabilities [1, 2]. It can be found in many modern mobile devices and cloud computing systems [3]. NN architecture, from its name, contains similar component as in human brain. The main two components are called Neuron and Synapse [1]. The neuron acts as the central brain, which accumulate charges and synapse is the connection between neuron to charge and discharge neuron [3].

There are different types of NN [1], e.g., Deep NN [4], Conventional NN [5], which used in image processing and Spiking NN [6]. Each type must be trained with any algorithm whether it is supervised (off-line) or unsupervised (online) algorithm. Online gradient descent algorithm [7] is one of the most algorithms used to learn the Multilayer Neural Network, MNN.

Memristor, the fourth passive circuit element discovered in [8] and fabricated in [9] is a two terminal electron device with nonlinear resistance performance. Its resistance changes according to the applied bias and maintain its final value before removing the bias [9]. This makes the device acts as a nonvolatile memory, and its memory applications can be found in [10-13].

The memristor has many advantages to be used in MNN. Its small size makes it very suitable for small algorithm circuit implementation. As in Online gradient descent algorithm, memristor carry the multiplication overhead of the weights and attribute of machine learning task [14,15]. This can be seen from its current-voltage relation [9]:

$$i(t) = G\big(s(t)\big)v(t) \tag{1}$$

Where G(s) is the conductivity which is function of the state (position) of the resistance of the device in real time [8].

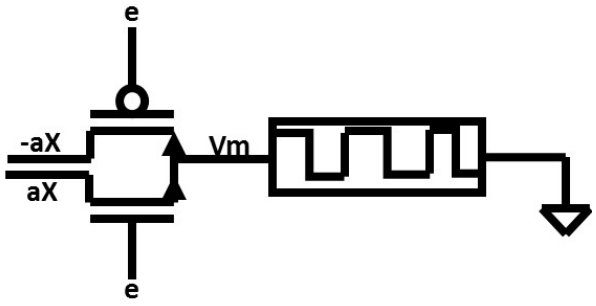

Fig. 1 The memristor-based synapse used in [16].

The nonlinear multiplication of the weights of the synapse can be achieved using (1). One of the most recent synapse implementations with memristor to achieve the advantage of this small size multiplication is in [16].

The rest of the paper is organized as following, the memristor-based synapse used in [16] for applying Online gradient descent algorithm in MNN, is introduced in section II. The proposed nondestructive reading mechanisms are illustrated in section III. A refreshment control circuit for the synapse in section II is proposed in section IV. Finally, a verification by simulation for contribution in section III and IV is shown in section V.

II. MEMRISTOR-BASED SYNAPSE

Fig. 1 shows the synapse used in [16], which used in scalable online learning algorithm in neural networks [7]. The synapse is a combination of one memristor for weight storing and two CMOS transistors for read and write control. The input of the transistors is the machine learning attribute, X, with scaling factor, a, to convert it to voltage units. The two transistors are chosen in a way that their conductivity is very high compared to the memristor one. In this way, the memristor input voltage is approximately equals the input of the transistors. To achieve this high conductivity, the scaling factor, a, is chosen to be in milli-volts and the control signal, e, is high enough to make the transistors in deep triode region of operation.

The control signal, e, is used to control the read and write of the synapse's weight into the memristor [16], See Fig. 2(a). The signal e has two phases, first phase is used for reading and second one for writing the updated weight. The reading is nondestructive, as it contains two out of phase input signals, Vm (aX and -aX) with same pulse width. In the write phase; the error, y in (2), which converted to time units by factor b, is controlling the duration of writing pulse width.

$$y = d - r \tag{2}$$

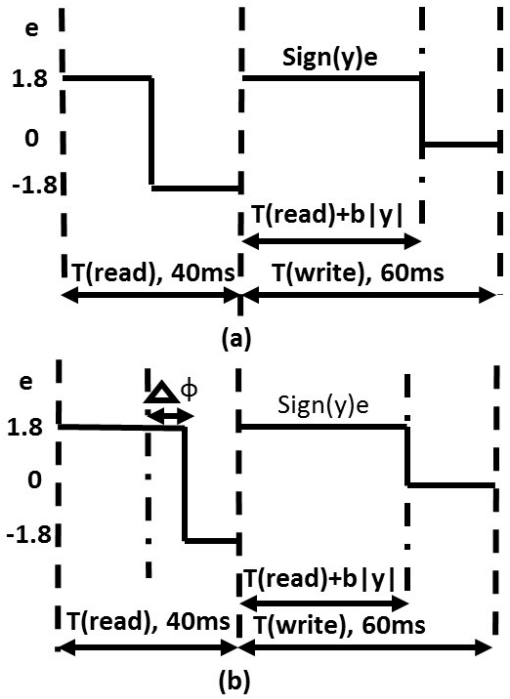

Fig. 2 The control signal e for both read and write of synapse weight. (a) original signal, and (b) with pulse width mismatch.

$$r = WX \tag{3}$$

Where d and r are the desired output and the computed output, respectively [16]. W is the synapse weight.

This synapse showed a good result in machine learning tasks simulation [16], however, it has a disadvantage in practical. The problem is that, making ideal pulse width for both negative and positive signals is impossible [17], see Fig. 2(b). This pulse width mismatch will corrupt the stored weight after multi periods of operation, which will highly affect the neuromorphic computing of the machine learning task (give wrong weights; wrong task evaluation). Specially in this synapse, the error will be very destructive, as the weights are changing in very small range (milli volts). The results of this error for different mismatch values is shown in section V.

III. Reading Techniques

In this section, two different reading signals are introduced. These two signals are less destructive compared with the original signal in Fig. 2. There corruption will be noticed after hundreds of periods, which rarely happens in machine learning tasks because of many trials (positive and negative) of attribute X [16] during computing. Both new signals are based on synapse circuit in the appendix of [16], which is less in size than their original synapse, see Fig. 3. However, the authors of [16] used their same destructive signal and increased the writing period, which will decrease the speed of the computing. In this paper, a less destructive reading signal and same original periods [16] of read and write are used, which is a great advantage of the proposed techniques.

Note that the term original refers to the synapse used in [16] and its signals. Also, the writing period is same for all signals proposed as in original synapse.

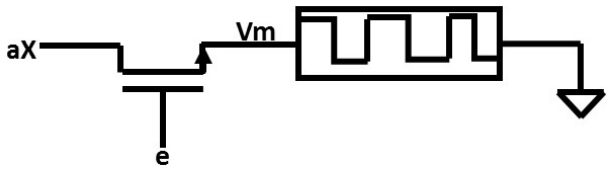

Fig. 3 Synapse Circuit in [16] appendix

A. *new signal inspired from [16,18]*

In Fig. 4, the new signal proposed is the same nondestructive concept used in [16] with the advantage of Nano scale as in [18]. Making the read pulses in Nano scale makes the pulse width mismatch effect, see Fig. 2(b), happens after several hundreds of periods. This is because the additional flux ΔØ will be very small and its accumulative effect will present after more periods than the one in milli scale. This signal is better than the second signal B in number of periods to fail as can be seen in section V.

Note that the original period of the reading phase is fixed in order to compare the results with the original synapse output.

B. *signal inspired from [18] only*

The same reading signal of memristor memory in [18] is used with same overall reading phase period for comparison with original signal, see Fig.5. In this signal, a very small pulse (1 Nano second) is applied to read the memristor state (weight). This will destruct the stored weight after approximately 310 period, as can be seen in section V. This technique is more destructive than signal in A, however, it is much better than the original signal.

IV. Refresh Control Circuit

It is possible to maintain the reading mechanism of the original synapse in [16] with making refreshment when failure happens [17]. The failure happens when the synapse is written with weight greater or equal to the weight of the next write period, which is without pulse width mismatch, see Fig. 6.

One possible approach for the refresh circuit is shown in Fig. 7. Here, the original synapse operats normally till failure happens, then a refresh full period (100 milli second) with same pulse width as in the write cycle of the original signal is applied. However, the amplitude of the pulse is negative in order to restore the weight to its original value, see section V. The pMOS is used with this interrupt pulse to modify the weight. Future work for this refreshment circuit is to reduce the interrupt period and even make the pulse used for weight restoring implemented in the original control circuit e.

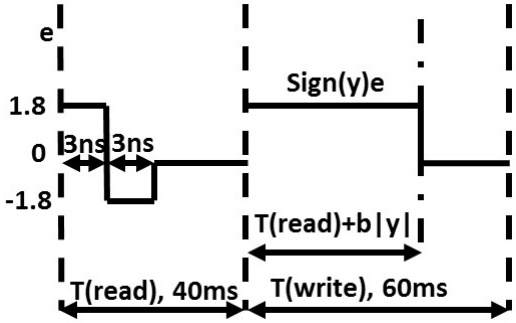

Fig. 4 The new signal inspired from [16,18]

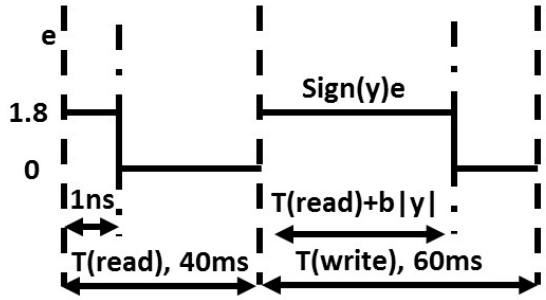

Fig. 5 Signal inspired from [18] only

V. Simulation Verification

The memristor Verilog-A Linear ion Drift model in [19] is used in Cadence Virtuoso. The CMOS 130nm technology is used for both transistors. All components and signals parameters used here and in [16] for good comparison are stated in Table I. T represents a full period length of both read and write.

Simulations of different pulse width mismatch, α, in the original signal were done and compared with original signal (without alpha, α) for producing the number of cycles to fail, see Table II. As can be seen, the relation between the number of cycles to fail and α is reciprocal, as it must be. Also, the range of cycles to fail is very small, which will highly destruct the stored weight.

Fig. 8 illustrates the difference between the new signal used in section III.A, with worst case pulse width mismatch α equals 10%, and the original signal (without α). As can be seen, there is no destruction for the stored weight even at 310T, which is a very high value for practical application [12]. This makes this new signal better than the other proposed one in section III.B. It is also inspired by the original signal mechanism.

In Fig. 9, the comparison between signal in section III.B and original signal is introduced. As can be seen, there is no destruction even at 310T, however, the destruction is more than signal in section III.A.

The refreshment circuit introduced in Fig. 7 is evaluated in Fig. 10 for α equals 5%. As can be seen, the control signal e is interrupted by a full period T with same pulse width as in write phase of original signal.

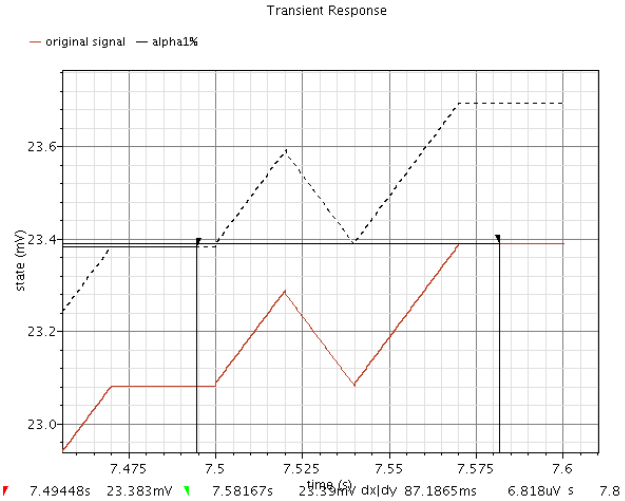

Fig. 6 The failure in writing the correct weight. Dashed signal is suffering from pulse width mismatch and solid one for original signal. Y-axis is the state signal of memristor (weight signal)

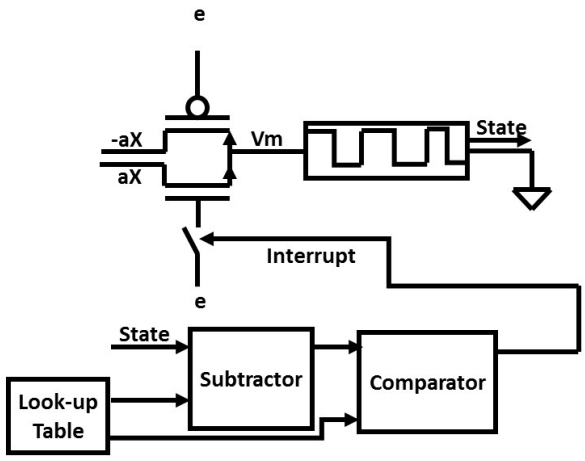

Fig. 7 The refresh control circuit block diagram.

TABLE I. Components Parameters

	Parameter	*Description*	*Value*	*Units*
Memristor	D	Length	10	nm
	Ron	Lowest Resistance	100	Ω
	Roff	Highest Resistance	100	kΩ
	μv	Linear ion Mobility	1e-14	$m^2/(s.V)$
Scaling	a	Input Scaling	1	mV
	b	Error timing conversion	0.6	T
Timing	T	Period	0.1	s
	Tread	Reading period	40	ms
	Twrite	Writing period	60	ms
Transistor	L	Length	130	nm
	W	Width	2	μm
Input	X	Data input	10	-
Control Signal	e	control signal amplitude	±1.8	V

TABLE II. Error Table

α	*No. of cycles to fail*
1%	76T
2%	39T
3%	26T
4%	20T
5%	16T

It is worth noting that the proposed refresh circuit can be used for different α. Note that from Fig. 8 to Fig. 10, the y-axis label (state (mV)) represents the weight in the memristor, and volts unit represent the place of the state (weight) in the memristor [19].

Acknowledgment

This work was partially funded by ONE Lab at Zewail City of Science and Technology and Cairo University, NTRA, ITIDA, ASRT, Mentor Graphics, NSERC.

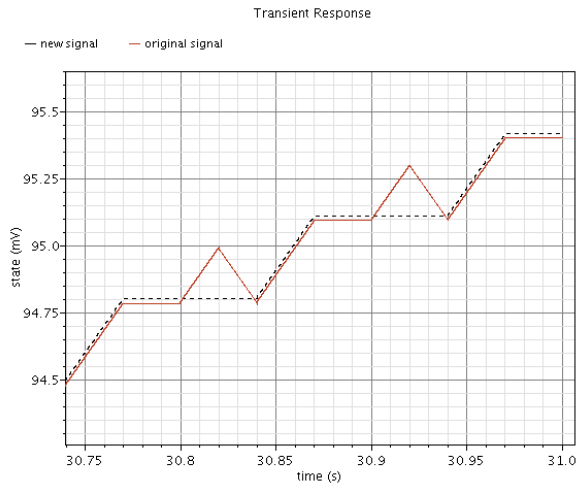

Fig. 8 Comparison of synapse weight by using signal proposed in section III.A with α=10% and original signal in [16]. As can be seen, the proposed signal has nondestructive effect on the synapse weight even after hundreds of T.

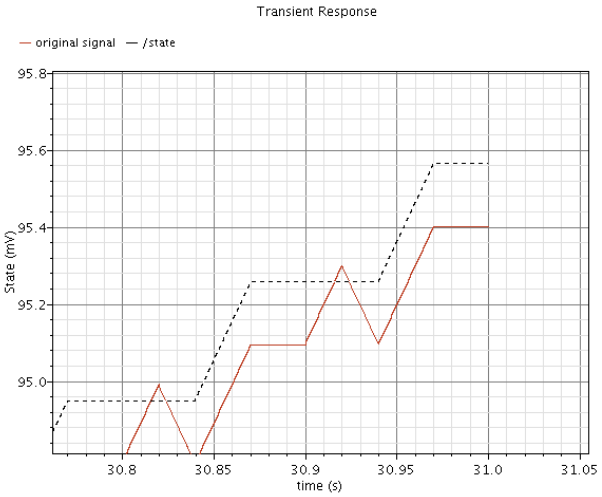

Fig. 9 Comparison of synapse weight by using signal proposed in section III.B and original signal in [16]. As can be seen, the proposed signal has destructive regime but after hundreds of T which is rarely happens in practical.

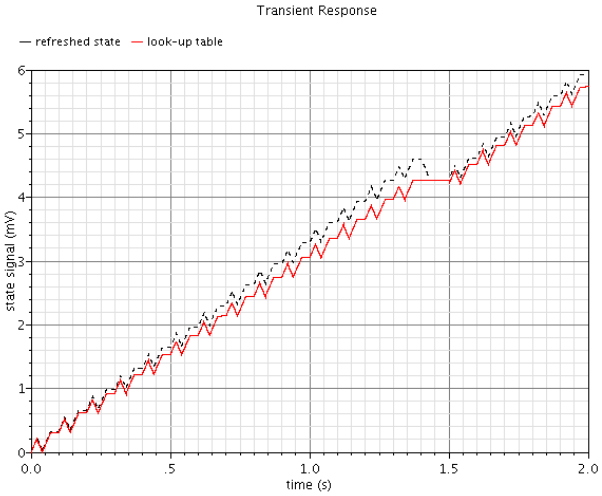

Fig. 10 The refreshment circuit for destructive reading happened with original signal with pulse width mismatch equals 5%.

REFERENCES

[1] O. Krestinskaya, A. P. James and L. O. Chua, "Neuromemristive circuits for Edge Computing: A Review," in *IEEE Transactions on Neural Networks and Learning Systems*. doi: 10.1109/TNNLS.2019.2899262.

[2] E. Adel, R. Magdy, S. Mohamed, M. Mamdouh, and H. Mostafa, "Accelerating Deep Neural Networks Using FPGA ", *IEEE International Conference on Microelectronics (ICM 2018)*, Sousse, Tunisia, pp. 180-183, 2018.

[3] C. Frenkel, M. Lefebvre, J. Legat and D. Bol, "A 0.086-mm^2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 1, pp. 145-158, Feb. 2019.

[4] Cheng, L. Xia, Z. Zhu, Y. Cai, Y. Xie, Y. Wang, and H. Yang, "Time: A training-in-memory architecture for memristor-based deep neural networks," in 2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC), June 2017, pp. 1–6.

[5] C. Yakopcic, M. Z. Alom, and T. M. Taha, "Memristor crossbar deep network implementation based on a convolutional neural network," in 2016 International Joint Conference on Neural Networks (IJCNN), July 2016, pp. 963–970.

[6] H. Mostafa, C. Mayr, and G. Indiveri, "Beyond spike-timing dependent plasticity in memristor crossbar arrays," in 2016 IEEE International Symposium on Circuits and Systems (ISCAS), May 2016, pp. 926–929.

[7] Bartlett, Peter & Hazan, Elad & Rakhlin, Alexander. (2007). Adaptive Online Gradient Descent.. Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference.

[8] L. O. Chua, "Memristor-the missing circuit element," *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507–519, Sep. 1971.

[9] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, Mar. 2008.

[10] Y. Ho, G. M. Huang and P. Li, "Nonvolatile memristor memory: Device characteristics and design implications," *2009 IEEE/ACM International Conference on Computer-Aided Design - Digest of Technical Papers*, San Jose, CA, 2009, pp. 485-490.

[11] H. Hossam, M. Dessouki, and H. Mostafa, "Time-Based Read Circuit for Multi-Bit Memristor Memories", IEEE International Conference on Modern Circuits and Systems Technologies (MOCAST 2018), Thessaloniki, Greece, pp. 1-4, 2018.

[12] H. Hossam, G. Mamdouh, H. H. Hussein, M. El-Dessouky, and H. Mostafa, "A New Read Circuit for Multi-Bit Memristor-Based Memories Based on Time to Digital Sensing Circuit", IEEE International Midwest Symposium on Circuits and Systems (MWSCAS 2018), Windsor, Ontario, Canada, pp. 1114-1117, 2018.

[13] A. Abubakr, A. Ibrahim, Y. Ismail, and H. Mostafa, "The Impact of Soft Errors on Memristor-Based Memory", IEEE International NEW Generation of Circuits and Systems (NGCAS 2017), Genova, Italy, pp. 229-232, 2017.

[14] L. Bottou and O. Bousquet, "The tradeoffs of large scale learning," in *Optimization for Machine Learning*, S. Sra, S. Nowozin, and S. J. Wright, Eds. Cambridge, MA, USA: MIT Press, 2011, p. 351.

[15] O. Krestinskaya, K. N. Salama and A. P. James, "Learning in memristive neural network architectures using analog backpropagation circuits," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 2, pp. 719-732, Feb. 2019.

[16] D. Soudry, D. Di Castro, A. Gal, A. Kolodny and S. Kvatinsky, "Memristor-Based Multilayer Neural Networks with online gradient descent training," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2408-2421, Oct. 2015.

[17] Y. Ho, G. M. Huang and P. Li, "Dynamical properties and design analysis for nonvolatile memristor memories," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 4, pp. 724-736, April 2011.

[18] M. Elshamy, H. Mostafa, Y. H. Ghallab and M. S. Said, "A novel nondestructive read/write circuit for memristor-based memory arrays," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 11, pp. 2648-2656, Nov. 2015.

[19] S. Kvatinsky, K. Talisveyberg, D. Fliter, A. Kolodny, U. C. Weiser and E. G. Friedman, "Models of memristors for SPICE simulations," *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, Eilat, 2012, pp. 1-5.

.